Using the Alberta Health and Wellness (AHW) lookup file to aggregate data to different geographic boundaries

Geographic Methodology Series No. 2

Prepared by
Health Surveillance
Alberta Health and Wellness
Edmonton, Alberta

Latest Revision: May 2004

Alberta
HEALTH AND WELLNESS

For more information contact:

Health Surveillance
Alberta Health and Wellness
P.O. Box 1360
10025 Jasper Avenue
Edmonton, Alberta
T5J 2P4

Phone: 780-422-4518
Fax: 780-427-1470

Series title: Alberta Health and Wellness Geographic Methodology Series
ISBN: 0-7785-2712-3 (online)
Geographic Methodology Series No. 2
Using the Alberta Health and Wellness (AHW) lookup file to aggregate data to different geographic boundaries
ISBN: 0-7785-3450-2 (pdf)

Note: Users of the information presented in this document in the analysis of health data must insure that products conform to the Alberta *Health Information Act*.

105-05430/

Executive Summary

This document is part of a family of reports that illustrates and documents the geographic methods required to properly analyze health data in Alberta. The descriptions and methods used are consistent across these reports. Together they provide all needed information required to properly understand the spatial component of health data.

This report documents the methodologies used to aggregate health, population, survey and other relevant data to different reporting boundaries. Data is usually collected at a finer geographic resolution than that used to report the data. In Alberta, for example, most of the health and population data is collected at the postal code level, but is presented at the RHA level. Since the data is available at the postal code level, then it is possible to present the data using other alternative reporting boundaries. Some of the reporting boundaries discussed include:

- RHA Dec 2003
- RHA March 2003
- RHA 2001
- RHA 1998
- RHA 1996
- RHA 1994
- Health Units
- Statistics Canada Census Boundaries
  - Enumeration Areas
  - Census SubDivisions
  - Consolidated Census SubDivisions
  - Census Divisions
- River Basins
- Geological Boundaries
- Provincial Electoral Districts
- Latitude Longitude Tiles
- Land Use
- SubRHAs
- Etc.

Acknowledgments

The following team prepared this report:

Alberta Health and Wellness

Erik Ellehoj	Geographic Information System Consultant
Dr. Donald Schopflocher	Biostatistician

The team acknowledges the contributions of the following reviewers in helping to insure completeness and accuracy in the report:

Larry Svenson	Alberta Health and Wellness
Tom de Beyer	Alberta Health and Wellness

The team acknowledges that some material was obtained from Statistics Canada, Canada Post, and Desktop Mapping Inc. web sites and summarized:

Table of Contents

Executive Summary 3

Acknowledgments 4

Table of Contents 5

I. Introduction 6

II. Cartographic Principles: Tessellation 6

III. Irregular Tessellation 7

- Irregular Tessellation: Regional Health Authorities (RHAs) 7
- Irregular Tessellation: Enumeration Areas 8
- Irregular Tessellation: Dissemination Areas 9
- Irregular Tessellation: Census Sub-Divisions 11
- Irregular Tessellation: Postal Codes 11
- Irregular Tessellation: Alberta Standard Geographic Codes 13
- Discussion 14

IV. Regular Tessellation 14

- Regular Tessellation: Latitude-Longitude Grid 14
- Regular Tessellation: Universal Transverse Mercator 17
- Regular Tessellation: Dominion Land Survey System 18
- Discussion 20

V. Using the Lookup File 21

- Dissemination Area Issues 23
- Linking Files 23
- Linking EA-DA information to Postal Codes 24

VI. Linking Legal (township-range) data to health event data 27

VII. Issues and Limitations 27

- Maintenance of the geographic conversion file 27
- Missing postal codes and invalid postal codes. 28
- Limitation of smaller geographic units 28
- Cross data-set linkage and conversion of boundaries 28
- Census blocks 28

VIII. Conclusion 28

References 30

I. Introduction

Analysis of health data usually implies the use of a set of geographic boundaries. These are used as a foundation in the analysis of any patterns that may be present in the data. Inconsistent sizes of mapping units create challenges for spatial research. Most studies use administrative boundaries, such as census divisions, census subdivisions, counties, and health regions. These boundaries may change over time and are may be inconsistent in size or shape. Alternative boundaries are numerous and their potential use is described in this document.

Health data may be aggregated to any type of geographic boundary. This procedure requires an understanding of how the boundaries may be created and how the health data is collected. It is possible to analyze health data using the appropriate lookup files once the procedure is well understood. AHW Surveillance has created lookup files that allow data analysis to several geographic boundaries.

This document describes the alternative types of geographic boundaries that may be encountered in health data analysis and methods that can be used to aggregate health event data to these alternative boundaries.

II. Cartographic Principles: Tessellation

For mapping purposes, the two most important characteristics of the earth are its size and shape. The size of any geographic object in a map can be estimated using a map scale. The shape of the earth introduces distortions on maps as the information is projected from a three-dimensional non-developable surface to a two-dimensional object (paper). Both issues influence how spatial information is presented or analyzed.

Tessellation or tiling is the process of partitioning continuous space. The decomposition of this continuous space is very well established. Continents, countries, states/provinces, counties, etc are all examples of partitioning continuous space. In Alberta, the province was divided into 17 discrete Regional Health Authorities in 1994. These boundaries were adjusted in 1996, 1998, 2001, March 2003, and December 2003 resulting in 9 regions at the time this report was written.

Any health or determinant data can be assembled based on these boundaries and presented in the form of tables, graphs, or maps. Tessellation, such as those found in RHAs, census boundaries, counties, provinces, are termed irregular as physical characteristics and political decisions determine their shape. Two accompanying documents: "Calculating Demographic and Epidemiological Quantities in Alberta by Geo-Political Areas Calculating Demographic and Epidemiological Quantities in Alberta by Geo-Political Areas" and "Definition of Sub-RHA Geographic Units in Alberta" describe how to work with these types of boundaries. These documents also explain how to summarize different data sources to these boundaries.

III. Irregular Tessellation

As was indicated above, irregular tessellation is the process through which a geographic entity is sub-divided into smaller units of inconsistent shape or size. Almost all administrative boundaries fit this description. Below, is a description of the more often-encountered irregular tessellation boundaries found in the Province of Alberta.

Irregular Tessellation: Regional Health Authorities (RHAs)

The province of Alberta was divided into 17 Regional Health Authorities (RHAs) by the provincial government in 1994. These boundaries were adjusted in 1996, 1998, 2001, March 2003, and December 2003 resulting in 9 regions at the time this report was written.. These authorities are responsible for the provision of health services to the citizens whose residences are part of the authority. To date, most analysis of health information in the province of Alberta has been performed using RHAs as the basic geographic reporting unit.

The RHA boundaries were generally formed from municipal boundaries and natural features. These boundaries were adjusted in 1996, 1998, 2001, and twice in 2003. RHA boundaries do not exactly match any other geo-political boundaries, so data collected using other smaller areas must be combined to calculate statistics for the RHAs. Figure 1 presents a map of the Regional Health Authorities, reflecting boundary changes introduced in December 2003.

Figure 1: Regional Health Authorities in Alberta, 2004

Irregular Tessellation: Enumeration Areas

Enumeration Areas (EA) are created by Statistics Canada for the purpose of the Federal Census performed every 5 years. An EA is the area canvassed by a single census taker and consists of approximately 200 homes, though some EAs may report fewer than 50 homes. Enumeration Area boundaries are determined after the census is completed and initial data has been tabulated. All population counts from the Federal Census are available at this level and it is the smallest standard reporting unit used by Statistics Canada up to, but not including the 2001 census. Within larger urban centres a more detailed geographic unit is available from a Street Network File (SNF) which contains single line streets, railways, and water features, along with names and address ranges. This unit, called a block face, refers to one side of a street between two consecutive intersections.

The EA boundaries and the EA centroid geographic coordinates assigned to each are available from Statistics Canada. The EA boundaries are formed using existing boundaries and natural features, whenever possible. Municipal boundaries established by the Provinces and rivers are the primary sources of boundaries. The EA boundaries are updated every five years when each population census is completed, and reflect any

changes made by provinces to municipal boundaries in the previous five-year period. A centroid is a single unique combination of latitude and longitude coordinates assigned to describe the location of a small area. In rural areas, EA centroids are generally at the approximate visual centres of the EA (as long as that centre doesn't fall inside a body of water). For those areas covered by the SNF, the EA centroids are based on the block face centroids.

EA centroids and block face centroids are generally used to assign geographic coordinates to postal codes.

Figure 2: Enumeration Areas (width = 100 km)

Irregular Tessellation: Dissemination Areas

In the 2001 census, the Dissemination Area (DA) replaced the EA as the basic unit of data dissemination. It is the smallest standard geographic area of which all census data are disseminated from the 2001 census onward. It was introduced because the EA populations had a great deal of variability and there were many EAs with very small populations. This required selective data suppression in order to maintain confidentiality.

The dissemination area is a small, relatively stable geographic unit composed of one or more blocks. The DAs respect the boundaries of the census subdivisions (CSD) and census tracts (CT). DAs therefore remain stable over time to the extent that CSDs and

CTs do. DAs are uniform in terms of population size (targeted at 400 to 700 persons) to avoid data suppression. However, some DAs may have lower (including 0) or higher populations in order to respect the boundaries of the CSDs or CTs. Note that the GIS boundary file for DAs contain more information about hydrography such as both sides of a large river and smaller lakes.

DA centroids are created in the same manner as EA centroids. DA centroids will be used in future assignments of postal codes to geographic coordinates.

Figure 3: Dissemination Areas (width = 100 km)

Irregular Tessellation: Census Sub-Divisions

Census subdivisions (CSDS) are created by Statistics Canada as geographic divisions larger than EAs and used for reporting of Census information. Since all census boundaries are coterminous, forming a hierarchy of geographic areas such that larger areas are formed as exact aggregates of smaller areas, all Enumeration Areas within a CSD are completely contained by the CSD and there are no partial regional assignments.

CSDs are formed using self-governing geographic entities. In Alberta this includes counties, Improvement Districts (IDs), Municipal Districts (MDs), Special Areas (SAs), towns, cities, villages, and Indian Reserves. They are formed using these existing municipal boundaries and are updated every five years as the municipal boundaries are revised (in an expanding city, for example). The boundaries are fixed on January 1st of the Census year.

Figure 4: Census Sub-Divisions (width = 100 km)

Irregular Tessellation: Postal Codes

Postal Codes are defined and maintained by Canada Post for the sorting and delivery of mail. Postal Codes are also widely used as a means of geo-coding databases for the purposes of location analysis, demographic analysis, and other types of geographical enabled analyses.

The structure of a Postal Code is fixed. The sequence is always Alphabetical character / Number / Alpha (full space) Number / Alpha / Number. The first three characters of the postal code are known as the Forward Sortation Area (FSA). The first character of a

postal code is allocated in alphabetic sequence from East to West across Canada and denotes a province, territory or a major sector found almost entirely within the boundaries of a province. (There are several instances where the first digit from the adjoining province is used for some residents or businesses of the other province. In Alberta this phenomenon is observed in Lloydminster). Rural postal codes can be distinguished from Urban postal codes by character "0" (zero) in the second position. The third character in an urban code identifies a postal station or city post office. In a rural code these characters identify a set of post offices in a geographic area.

The last three characters - known as the Local Delivery Unit (LDU), will guide the mail to a specific location. In an urban area, the LDU refers to a few buildings along a street, or to a group of mailboxes in suburban areas, or to a wide area (spaghetti-like route) served by a rural route service from an urban post office. For a rural destination, the LDU refers to the community to which the mail is to be delivered. Postal communities (defined by Canada Post) usually do not respect municipal boundaries (defined by the provinces). The letters W and Z are not used as the first letters of postal codes; D, F, I, O, Q and U are never used in Canadian postal codes. If one of these letters appears in your records, it is inaccurate and should be corrected.

In new suburban areas, postal codes may be linked to a community mailbox. These boxes can service both odd and even sides of the same street, or different streets, within a 300 metre radius of the community mailbox. Often, rural postal codes represent the location of the place where the mail is sorted and not the final place of delivery. This can mean that one postal code, with multiple positions, may be used to provide service to several small rural towns or parts thereof. Most rural postal codes provide service to parts of more than one legal municipality. Boundaries are available for Forward Sortation Areas (FSAs). FSAs are small in urban areas. There are between 20 to 30 FSAs in cities of 750,000 and 2 or 3 in cities between 50,000 and 75,000. FSAs are very large in rural areas and so are of limited use for geographic coding in these circumstances.

Each postal code is defined by a set of postal delivery locations, which are listed as address ranges. Boundaries for individual (6 digit) postal codes are not available (Canada Post does not need them for their own purposes). It is difficult to generate boundaries for the nearly random set of dots indicating the location of persons using a set of post office boxes with the same postal code. The challenge only increases when dealing with the spaghetti-like routes from urban post offices, spilling into the map territory of the adjacent rural FSAs.

Canada Post updates postal codes monthly. This creates challenges in maintaining the implied boundaries (or geographic correspondences) for every postal code. Most of the new postal codes created in Alberta are assigned to expanding urban areas

Figure 5: Postal Code Centroids and CSDs (width = 100 km)

Irregular Tessellation: Alberta Standard Geographic Codes

In Alberta, a unique code is created and maintained by Alberta Government Services and assigned by Alberta Vital Statistics to any community where a birth, death, or marriage took place.

While the term Standard Geographic Code abbreviated as SGC is sometimes used for these units, this has caused great confusion because Statistics Canada uses SGC as an alternate term to refer to CSD. These codings are not the same, and the term Alberta Standard Geographic Code (ASGC) will be used here for the coding used by Alberta Vital Statistics. ASGCs were originally loosely based on Statistics Canada's CSDs. The first digit of the CSD was removed and an additional digit was appended to distinguish the rural and urban component of a community. (If there is insufficient information to make this determination, then the ASGC appropriate to the urban component is assigned).

In assigning an ASGC code, postal code and street address are disregarded and only community name is used. ASGCs designate towns, villages, cities, and some reserves. Since their creation ASGCs have been assigned to areas that do not have a corresponding CSD assignment. These include smaller villages, hamlets, and communities that no longer exist.

In those instances where a community name does not have an ASGC, Alberta Government Services performs a manual process to identify the community. If required, a new ASGC code is created. Approximately 15 new ASGCs are created each month, most

of them for historical purposes. For example, several individuals were born in Dogpound (a small community near Madden) but this community no longer exists. An ASGC code is needed in order to record it as a birthplace for all the appropriate individuals as their deaths are recorded.

Individuals who were born or died outside the province are assigned a code according to their circumstances. Unique codes are used to identify each province and country.

Further information about geographic boundaries can be obtained from the respective agencies, and most of this information is available from the web. There is a considerably larger set of available boundaries if the list of agencies is expanded beyond Alberta Health and Wellness and Statistics Canada. These are not relevant and are not discussed here.

No boundaries or point file exist for ASGCs and therefore no map can be presented here.

Discussion

The tessellation methods described above are typically found when working with health data in the province of Alberta. The boundaries used for health data in other provinces or countries are different, but the basic methods of presenting the data are similar. Likewise, working with other provincial jurisdictions, such as Agriculture, AADAC, justice, etc. implies the use of different geographic boundaries specific to the particular jurisdiction.

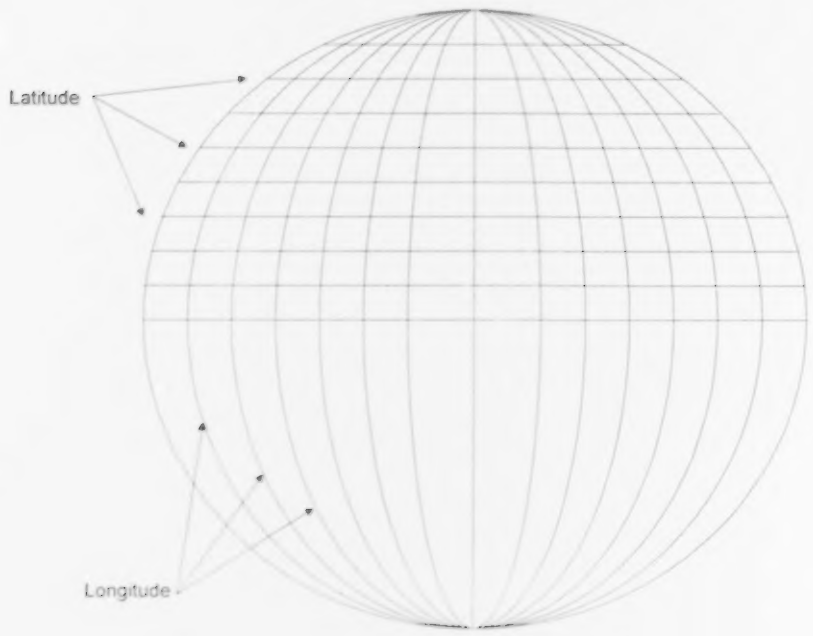
IV. Regular Tessellation

An alternative form of discretization is the use of a regular tessellation. Examples of regular tessellation are the organization of map sheets in a topographic mapping series, image data from remote sensing, organization of map libraries, and uniform sampling of a continuous spatial distribution. These examples are all based on a rectangular shape, but alternative shapes may be used for regular tesselation, such as triangles and hexagons. If multiple resolutions of the tiles are to be offered, then rectangles and triangles offer greater simplicity. In this report, only rectangular shapes are explored.

Regular Tessellation: Latitude-Longitude Grid

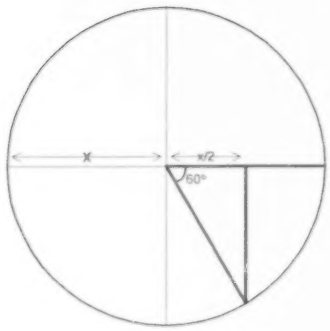
A common method used to perform uniform tessellation involves the use of latitude-longitude lines. Lines of latitude start at the equator or 0 degrees and increase toward the poles to 90 degrees North and South. Lines of longitude form great circles that divide the earth into two equal halves and cross both poles. The line of longitude that crosses Greenwich has been assigned the value of 0. Values increase to the east and west until the reach a maximum of 180 degrees at the date line. All lines of latitude are parallel while lines of longitude are sometimes termed "meridians".

Figure 6: Latitude and Longitude



Although the distance between any two lines of latitude is always the same regardless of longitude the same is not true for distances among lines of longitude. All lines of longitude converge at the pole and therefore, at that point, the distance between any two lines of longitude is 0. For example, the distance represented by one degree of latitude is approximately 111 km (69 miles or 60 nautical miles). The distance represented by one degree of longitude is approximately 111 km x cos (latitude). A degree of longitude represents half the physical distance at 60 degrees of latitude than at the equator (cosine of 60 = 0.5).

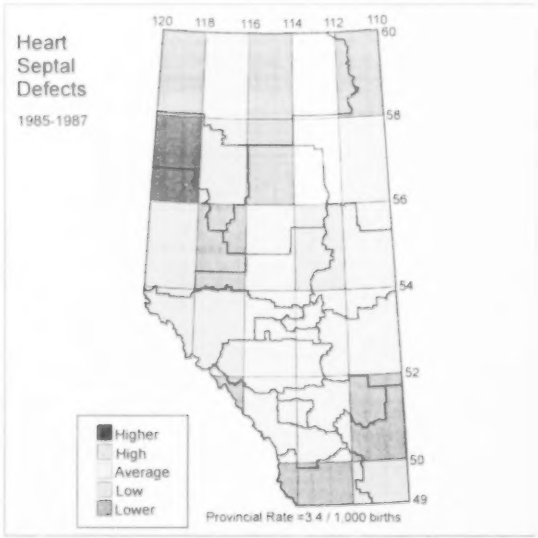
Figure 7: Distance at 60°.



Geographic tiles constructed with lines of latitude and longitude are common, despite the gradual shrinking of the size of tiles as proximity to the poles increases. The net impact of the technique for most of the populated world is that some geographic tiles are half the size of others. Despite the change in size according to proximity to the poles, it is quite easy to visualize these tiles at a local or global level.

The province of Alberta is well suited to a tessellation based on longitude and latitude lines. The northern and southern boundaries are formed by latitude lines. The eastern boundary is formed by a longitude line, and half the western boundary is formed by a longitude line.

Figure 8: 2 degree by 2 degree latitude longitude tiles



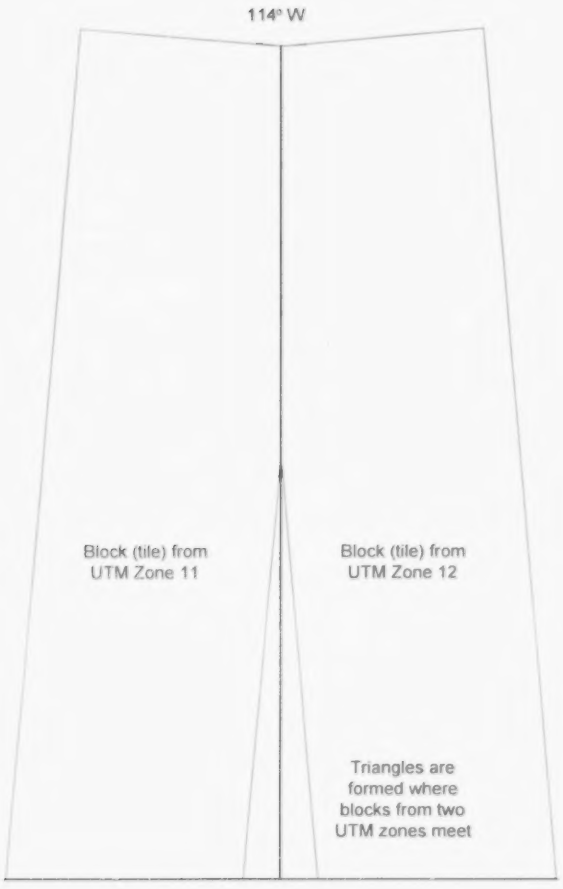
Regular Tessellation: Universal Transverse Mercator

Universal Transverse Mercator (UTM) tiles have also been used to define consistent geographic regions. Explanations about the UTM grid can be found in many locations including: http://www.tpub.com/inteng/9h.htm. The Universal Transverse Mercator (UTM) is an international coordinate system developed by the U.S. Army. The UTM divides the world into 60 zones of 6 degrees longitude. These zones are numbered consecutively beginning with Zone 1, between 180° and 174° west longitude, and progressing eastward to Zone 60, between 174° and 180° east longitude. In each zone, coordinates are measured north and east in meters. The northing values are measured continuously from zero at the Equator, in a northerly direction. Grid values to the west of this central meridian are less than 500,000, to the east more than 500,000. A value of 0 or 1,000,000 meters is never reached within a UTM zone.

The challenge associated with the use of tiles involves consistency, visibility and understanding the basis of the system. Where two UTM zones meet, they form triangular-shaped tiles. These tiles are not of the same size as regular UTM tiles, and may be too small to be visible. In addition, the UTM coordinates are not as widely understood as longitude and latitude coordinates. These factors suggest that, for epidemiological surveillance purposes, the longitude-latitude tile method is a better tool than the UTM tile, although both are valid alternatives.

The Universal Transverse Mercator (UTM) tiles have been used to define consistent geographic regions globally. Several bird atlases use these tiles as a foundation, such as "The Atlas of Breeding Birds of Alberta" [1] and "Atlas of the Breeding Birds of Ontario"[2]. In Canada, the UTM grid appears in all 1:250,000 and 1:50,000 scale maps.

Figure 9: Triangles formed by tiles from two UTM zones



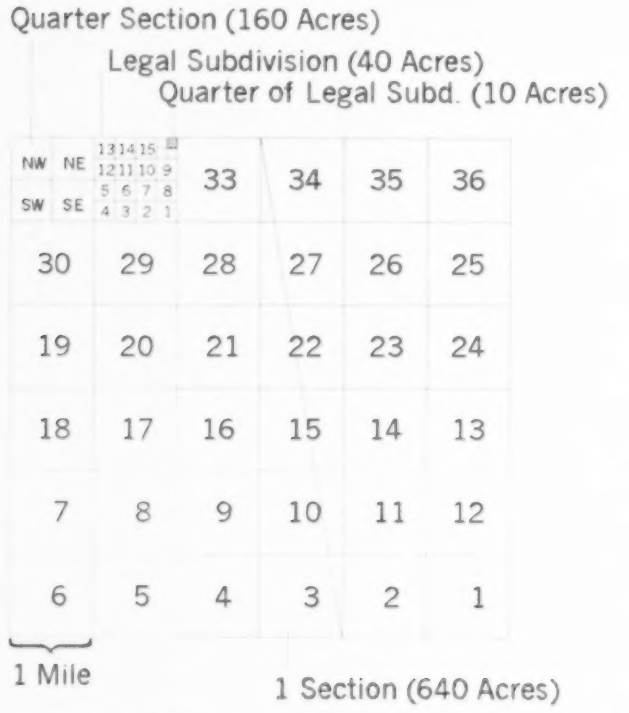
Regular Tessellation: Dominion Land Survey System

The Dominion Land Survey System used in Western Canada divides the land into equal units of six miles by six miles (approximately 10 km x 10 km). These units are further subdivided into 1 mile by 1 mile "Sections". These are further sub-divided into quarter sections or Legal Land Subdivisions. A correction line appears 12 miles from the Canada-US border (49° N) and at 24 miles intervals, thereafter. There are three reference meridians in Alberta, the fourth meridian forms the Alberta-Saskatchewan border, the fifth meridian (at 114° W) divides the City of Calgary, and a third meridian appears at 118°. Rural roads follows the township and range boundaries.

The theoretical boundaries of each geographic unit (such as township, quarter section) can be estimated quite accurately, but in practice, the physical location of the boundaries are inconsistent with these theoretical boundaries. A detailed survey of the entire

province has not been completed for the township grid. therefore parts of the province rely on a theoretical grid, and other parts rely on an actual grid (with associated errors).

Figure 10: Township subdivisions



1 Township = 6 Miles by 6 Miles (36 Square Miles)

Most of the health data available in Alberta has a coarser geographic resolution than the township level, therefore this grid is currently too fine for analysis of health data. Smaller land units at the furthest western edge of each meridian create difficulty in the display of any geographic information with this method. This system serves the needs of the agricultural, forestry and the oil and gas industries. Some natural history work has also been done with this grid, namely the Atlas of Alberta Ungulates.

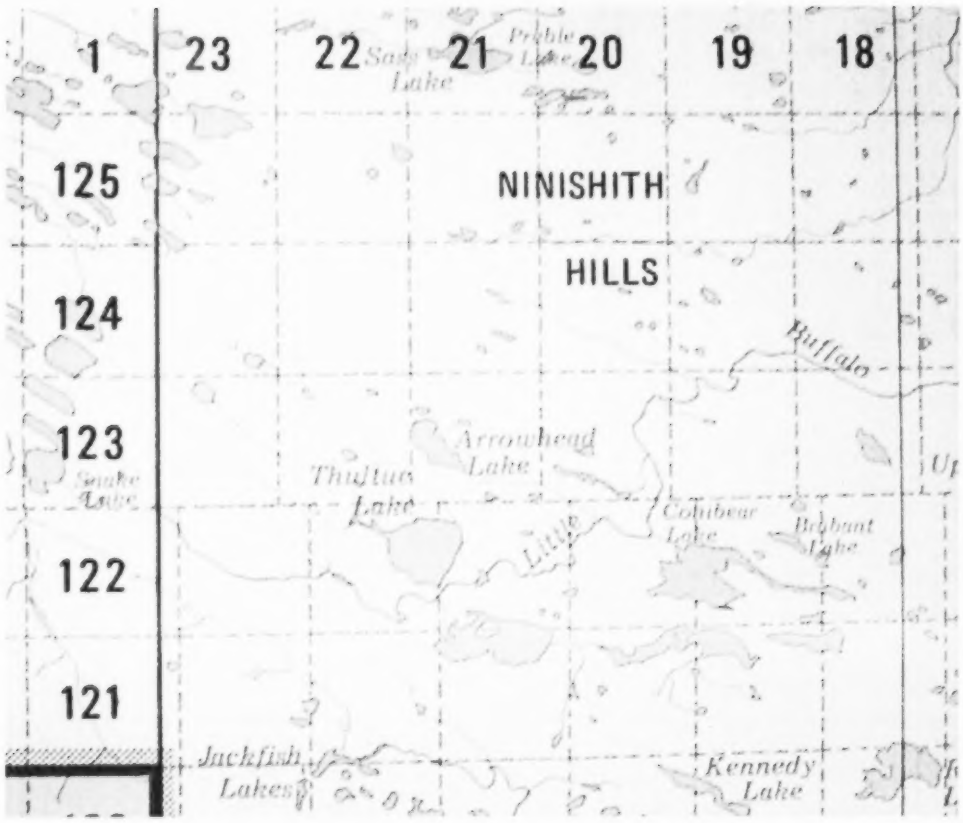


Figure 11: Partial townships at western edge of the fifth meridian

Discussion

The section of the map below, part of the 1:250,000 series, illustrates all three regular tessellation methods. Latitude-longitude lines define the map sheet boundaries. The township grid is identified using grey lines. The township numbers appear on the left side (Tp) and ranges on the top (R). The UTM grid appears in blue lines and coordinates appear on the edges of the map sheet in blue numbers. The block designations appear on the right side where four corners meet (MC, NC, MB, and NB).

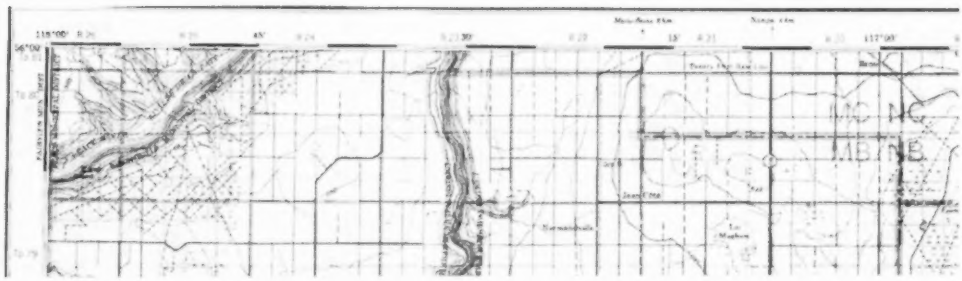


Figure 12: Portion of map sheet showing three regular tessellation methods

V. Using the Lookup File

In Alberta, much of the health data are available at the postal code level. In order to locate and assign health events/determinants into an appropriate geographic unit a master geographic data conversion file is used. This is typically accomplished using a postal code conversion file, a file that lists the geographic location of every postal code. Alberta Health and Wellness uses the Postal Code Conversion File created and maintained by Alberta Treasury. Updated monthly by Alberta Treasury, this file contains every postal code in the province and the Enumeration Area (EA) that best represents that population. Each postal code is assigned a single unique EA, and geographic coordinates (but not boundaries) can thus be generated for the postal codes based on the coordinates of the assigned EA centroids. A full description of the methodology appears in: "Calculating Demographic and Epidemiological Quantities in Alberta by Geo-Political Areas Calculating Demographic and Epidemiological Quantities in Alberta by Geo-Political Area.

Alberta Health and Wellness Surveillance have created an enhanced version of this file. The geographic relationship of all EA centroids in the province with respect to the suitable reporting boundaries, such as Health Regions, Census Divisions, historic health region boundaries, watersheds, and longitude-latitude tiles appear in the file. The table can be expanded to include other tessellation boundaries such as the township grid, UTM blocks, climatic regions, etc. Health data, collected at the postal code level, can thus be analyzed in context of any reporting boundaries that appear in this look-up file.

The most current version of the PCCF is stored in M:\AHW\PopulationandGeospatialInfo\PostalCodes\Post Code.mdb in an MS Access database. The
database contains the following tables:

CCSD96
This table contains a summary of the Consolidated Census Sub-Division (CCSD) information as listed by Statistics Canada. For each CCSD, the table lists corresponding Census Division (CD), CCSD Name, Area of the CCSD, CCSD Population, and the

number of dwellings in the CCSD. For a description of CCSD please refer to the Statistics Canada web site, or to the document entitled "Calculating Demographic and Epidemiological Quantities in Alberta by Geo-Political Areas".

CD96
This table contains the information for each Census Division. The table only lists CD number and name.

CSD Types
This table is a look-up list whose parent is CSD96. The table lists an explanation for the field CSD type that appears in the table CSD96.

CSD96
This table lists every CSD in the province of Alberta. The Census Division number, CSD Number, and CSD Name all appear in the list. CSD type is used to determine if the CSD is a City, County, Reserve, ID, etc. Further information about the CSD types appears in the table named CSD Types. Other information available are the areas, population, and number of dwellings for each CSD and corresponding CCSD.

FedEA96
This table is very similar to CSD96, but it works at a greater level of detail. The Enumeration Area is the smallest geographic data unit for which data is currently available from Statistics Canada. For each EA it lists the corresponding CSD, population and number of dwellings. It also lists a unique pair of latitude/longitude coordinates for each EA which is usually termed an EA centroid. This pair of coordinates identifies a location within the EA polygon, but not necessarily in the centre. These coordinates are later used to assign locations to postal codes.

PCFedEA96
This file contains the EA assigned to each postal code. The Alberta Treasury version of the PCCF lists only one EA for each postal code although the postal code may serve several EAs. For a discussion about this issue, please consult the document entitled "Calculating Demographic and Epidemiological Quantities in Alberta by Geo-Political Areas".

The file also lists the Municipality name, retirement status (Active or Inactive), Population Density assignment, postal code implementation and/or retirement date. This file is very important because it allows assignment of information known about the EA to each postal code and it is through this methodology that coordinates are assigned to postal codes.

Ped_by_PC
This file contains a list of the postal codes and the appropriate Provincial Electoral District (PED). The names of each PED can be obtained by linking to the file called PEDNames.

RHA
RHA contains the RHA number and the corresponding RHA name.

RHAData96
This file contains the RHA assignment for each EA. Note: do not use this table because the RHA assignments are based on the 1996 boundaries, not the current boundaries that were established in 1998.

RHAover Time
This file contains every postal code (current and historic) and for each it lists RHA assignment based on current (1998) and historic (1996 and 1994) boundaries. It also lists the municipality name and Provincial Electoral District (PED).

Dissemination Area Issues

The 2001 census introduced the Dissemination Area (DA) as a basic geographic unit. The agencies responsible for creating and maintaining the postal code lookup files decided to undertake the lengthy procedure in order to ensure that all their files used DAs as a basic geographic unit instead of EAs. A DA lookup file has been created and it contains an identical set of fields as those found in the EA lookup file. The new PCTF (2004) contains EA and DA assignments for every postal code. A small amount of discrepancy is expected when comparing EA to DA summaries since the centriods for EAs and DAs are not always identical. The total size of this discrepancy has not yet been evaluated. An analysis of DA to EA discrepancies will be examined at the same time as an analysis of block assignments vs EA & DA assignments. The analysis will be performed on a variety of small area boundaries to determine the impact in the worst case scenarios.

Linking Files

The last table is the most useful for most people because it allows for the generalization of all data available at the postal code level to be generated at the RHA level. Many people only work at the RHA level and this file provides all the necessary tools to generate the needed information. If an individual needed to work with data at the postal code to show cases in the province, or to aggregate data to a different set of geographic boundaries, then this file can't provide the needed information. Historically a file had been generated to provide information about the assignment of the postal codes to alternative geographic regions such as Latitude Longitude Blocks, or hydrographic basins. This method required regular checks on the postal code lookup file to ensure that it wasn't newer than the extra look-up file. A new method is now used to obtain this information and it requires an understanding of how the PCCF is generated:

For each postal code, the best candidate EA is listed through a well-established process at Alberta Treasury. The file that Alberta Treasury generates appears as table PCFedEA96 in the Access database. The table called fedea96 lists information about each EA as

Statistics Canada delivers it. The key pieces of information from a geographic point of view are the latitude and longitude coordinates assigned to each EA. By linking the table RHAover Time to the table fedea96 using the EA code as a link, it is possible to derive geographic coordinates for each postal code. This extra information about the postal codes allow a wide range of geographic operations within a Geographic Information System (GIS). If the operations are to be performed on items that do not change on a regular basis, then it is possible to generate a list of geographic correspondence for each EA instead of performing the task with the postal codes that change every month. For example, the Latitude Longitude Block boundaries do not change and therefore any assignments to these are not likely to change unless the locations of the assigned objects change. If the assignment of postal codes were to be generated to the Latitude Longitude blocks, then this list would require updating every month as new postal codes are added to the list and others are retired. The assignment is performed using the geographic coordinates derived from the EAs as explained earlier. The same answers can be obtained if the assignments were to be done to the EAs. The boundaries and coordinates of the EAs change only every five years and therefore they are in a very stable geographic assignment. If the EAs are assigned to Latitude Longitude blocks, then this assignment will not change for a five-year period. Each postal code has been assigned an EA, and though this process any other information that has been assigned to the EA can also be assigned to the postal code. As new postal codes are added, these too are assigned to an EA, and therefore they are assigned a Latitude Longitude block or any other information associated with the EA.

Please note that the process is identical with the DA lookup file, but DAs are used instead of EAs for linking the files.

Linking EA-DA information to Postal Codes

The table called EAOL.dbf (Enumeration Area Overlay) that appears in this directory (M:\HealthSurveillance\Methods\Postal Code\Public) contains further assignments to the EAs. This file contains all the Enumeration Areas as listed in the 1996 Census of Canada. Enumeration Areas were used for this file as these are the geography founding block of the Postal Code lookup file used in Alberta Health (and in The Government of Alberta in general). A Dissemination Areas (DA) lookup file has also been created with an identical structure and it appears under the name of DAOL.dbf. The DAOL file should be used to examine any health data in context of the Dissemination Areas.

The postal code lookup file is updated monthly, therefore it is time-consuming to add all the fields present in the EAOL file to the postal code lookup file since the process is done monthly, but not at set dates.

The user must link the latest postal code lookup file using the field FedEA to determine the assignment of postal codes to the layers of information that are part of EAOL.

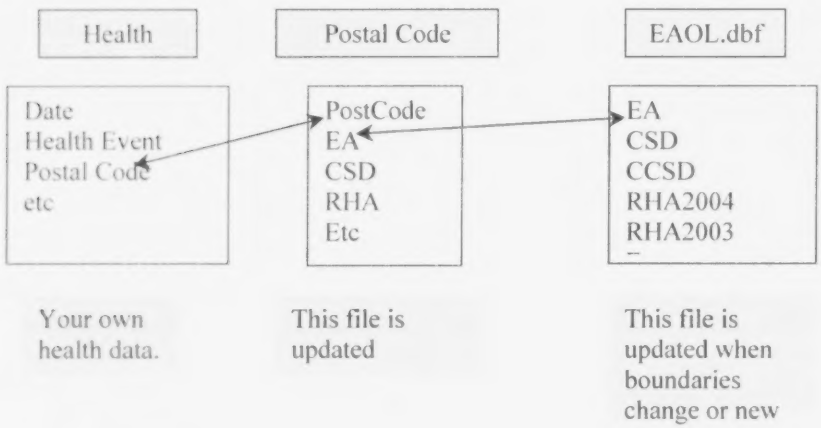
The fields in the file are:

- FedEA Enumeration area as it appears in the 1996 Census of Canada. Note that two leading digits are missing, namely "48" as they simply denote the Province of Alberta. The Postal Code lookup file makes the same omission, and this was the decision for eliminating the two digits. Please add "48" to all EA codes before linking with census data. The DAOL file uses a field called FedDA and is used in the same manner as FedEA.
- CSD Census Subdivision as it appears in the 1996 Census of Canada. Note that two leading digits are missing, namely "48" as they simply denote the Province of Alberta. A Census SubDivision is any self-governing entity such as County, town, village, city, MD, ID, reserve, and Special Area. Note that Metis Settlements are not CSDs despite their self-governing status since they only appear in Alberta.
- FEDEAPOP This is the total population for the EA as it appears in the 1996 census. FedDAPop conatains the DA population from the 2001 census in the DAOL file.
- FEDEAIND This is the urban status index that appears in the census and identifies urban core, urban perimeter, rural perimeter, rural, etc
- FedEADwell This is the total number of dwellings as it appears in the 1996 census. FedDADwell contains the same information in DAOL, but using DA 2001 census information.
- Long Longitude of the EA centroid (DA in DAOL)
- Lat Latitude of the EA centroid (DA in DAOL)
- TwpRge Township and range. The first character contains the reference meridian (4, 5, or 6 in Alberta) , the next two characters contain the range, and the last three characters contain the township number.
- GHD General Hospital Districts were created approximately 10 years and they represent the "trade area" or "catchment area" for the facilities that existed at that time
- GHD_2 This is almost identical to the field above, but some GHDs have been amalgamated and others split to reflect the facilities that maternity programs
- NRN Natural Regions (general category i.e. Boreal Forest)
- NSN Natural Regions providing greater specificity (dry mixedwood)
- EDN Name of the geographic region (i.e. Athabasca Plain)
- ERA Geology –Geological Era (i.e. Mesozoic)
- Period Geology –Geological Period (i.e. Cretaceous)
- Epoch Geology –Geological Epoch (i.e. late)
- Formation Geology –Geological Formation (i.e. Belly River Group)
- SUBRHA04 Sub-RHA boundaries effective Dec. 2003
- RHA2004 Regional Health Authorities, effective Dec. 2003
- RHA2003 Regional Health Authorities, effective 2003
- RHA2001 Regional Health Authorities, effective 2001
- RHA1998 Regional Health Authorities, effective 1998
- RHA1996 Regional Health Authorities, effective 1996
- RHA1994 Regional Health Authorities, effective 1994
- HUID Health Unit I.D. (used before regionalization)

- HUName Health Unit Name (used before regionalization)
- LandUse Land Use generated from satellite imagery. Minimum resolution cells of 1kmx1km
- LandUseCo Land Use codes (1 to 14) corresponding to the Land Use Names
- Basin River basin. The file AllBasin.dbf contains the rivers and streams in each basin (more than one water body is often listed for each basin). The field "Watershed" in Allbasin.dbf is used to link with this field.
- SubRHA_N Numeric field containing sub-RHA category (first character is the RHA ID)
- SubRHA_C Character field containing sub-RHA category, eff. Dec. 2003

This linking process allows for geographic analysis without the need to learn or use a GIS. The linking process is summarized by figure 12 (below).

Figure 13: Linking health-event data to the EA lookup file



VI. Linking Legal (township-range) data to health event data

Legal land description (aka township-range) locational information is available from a number of sources. Since this refers to property locations in rural areas, there is no direct method to link this information to postal codes. Most rural land owners use meridian, township, range, and section to identify their residence. The agriculture, forestry, and oil/gas industries also use this method to identify the location of their respective activities. A file was created to link these types of activities with health event data. This file is called TWP2PC2.dbf and it contains the following fields:

CODE	Meridian, Township, and Range merged into a single field
MERIDIAN	Meridian, in Alberta the only options are 4, 5, and 6
Range	Range or column west of the reference meridian
Township	Township or row starting at 49th parallel
Lat	Latitude of the township-range-meridian centroid
Lon	Longitude of the township-range-meridian centroid
Voro	A numeric field to identify postal code region
PC	Best-match (estimated) postal code for the twp-rge-mer.

In order to link legal land description with health data it will be necessary to create a new field in the table with legal land descriptions. This field must match the structure of the first field in twp2pc2.dbf by amalgamating meridian, township, and range into a single numerical field. For example the entry "401001" refers to a location west of the fourth meridian "4", range 1 "01", and township 1 "001". Once the field has been created it can be linked to any table that contains postal code information using twp2pc2.dbf. Legal land description data can also be aggregated to any of the boundaries present in the EAOL.dbf table by linking the legal land description data to the twp2pc2.dbf, and then to the postal code lookup file (to obtain EAs) and finally to EAOL.dbf using EA as the linking field.

There is no need for a mechanism to link at a finer level than the township-range-meridian (i.e. section, quarter-section, legal sub-division, etc) since the postal code areas cover areas much larger than the area 36 square miles comprised by the current methodology. A township is 6 miles high and a range is 6 miles wide.

VII. Issues and Limitations

Maintenance of the geographic conversion file.

Some postal codes may retire and some may emerge due to changes in populations and other reasons. Thus, updating the conversion file on a regular basis is essential to ensure the validity and timeliness of application of this method. The method used by Alberta Health Surveillance ensures that lookup files are current for a period of at least five years. This issue is true regardless of tessellation method used (regular or irregular).

Missing postal codes and invalid postal codes.

Missing data and data entry error are common in administrative databases. In Alberta, a small proportion of records did not have a usable postal code or did not have postal code at all during 1985-1998. This proportion varied by data sets and year. For example, less than 0.03 percent for the AHCIP stakeholder registry displayed this problem. While 3-12 percent for Hospital Morbidity, Physician Claims and Vital Statistics Birth Registry represented unusable postal code information. During 1991-1998, the missing and invalid postal codes maintain a fairly low level (less than 7%) for the majority of these databases. For those without a valid postal code, effort is required to assign the location by using other methods. The AHCIP Stakeholder Registry is a good source of geographic location because only a very small proportion of postal codes is missing or invalid.

Limitation of smaller geographic units.

Although the size of a geographic reporting unit tile can be made very small, the postal codes remain the unit of data collection. Rural area postal codes are large geographic units and therefore smaller-than-postal-code geographic units are not useful until health data is collected with greater geographic precision.

Cross data-set linkage and conversion of boundaries.

Starting from the smallest geographic unit is often desired in data linkage across spatial data files. Postal code is the smallest unit available in many health data systems in Alberta. This may not be always available in some data systems, such as Vital Statistics which uses Alberta Geographic Codes which are not codetermenous with postal codes. Please refer to "Calculating Demographic and Epidemiological Quantities in Alberta by Geo-Political Areas" for a full explanation of this issue.

Census blocks

Please note that lookup files are normally constructed using census blocks coordinates since they are the smallest boundaries available. The Statistics Canada PCCF uses blocks to assign geographic locations to every postal code. The EA and DA lookup files represent a generalized version of the more detailed blocks lookup file. In Alberta, the PCTF is more accurate and more current than the PCCF despite the fact that blocks offer greater geographic resolution. DA and EA lookup files were created in order to match the same level of resolution that is available from the PCTF. A more detailed file will be created if the PCTF adds census block assignments at some future time.

VIII. Conclusion

This report outlines the methods needed to aggregate data for health data analysis. This document is part of a family of documents that illustrates and documents the geographic methods required to properly analyze such data. These reports appear in the references section.

The descriptions and methods used are consistent among these reports and provide all needed information required to properly understand the spatial component of health data.

References

Ellehoj, E.A., Schopflocher, D, et al (2004). *Calculating Demographic and Epidemiological Quantities in Alberta by Geo-Political Areas Boundaries-Geographic Methodology Series No. 1.*

Ellehoj, E.A., Schopflocher, D, et al (2004). *Health Rate Mapping Template-Geographic Methodology Series No. 3.*

Ellehoj, E.A., Schopflocher, D, et al (2004). *Calculating Distances in Alberta-Geographic Methodology Series No. 4.*

Ellehoj, E.A., Wilson, S, et al (2005). *Definition of Rural and Urban Areas for Health Services Provision-Geographic Methodology Series (in preparation).*

Ellehoj, E.A., Schopflocher, D, et al (2005). *Geographic Activities in AH&W - Geographic Methodology Series (in preparation)..*

Health Trends in Alberta, 2001. Alberta Health Surveillance, Alberta Health and Wellness. ISSN 1480-6657